An Anatomy of a Lie:

Discourse Patterns in Customer Complaints Deception Dataset

Dina Pisarevskaya
Institute for Systems Analysis
FRC CSC RAS
Moscow, Russia,
dinabpr@gmail.com

Boris Galitsky
Oracle Corp
Redwood Shores CA, USA,
boris.galitsky@oracle.com

Jay Taylor
Oracle Corp
Redwood Shores CA, USA,
jay.taylor@oracle.com

Andrey Ozerov
Oracle Corp
Redwood Shores CA, USA, andrey.ozerov@oracle.com

ABSTRACT

Automated detection of text with misrepresentations such as fake reviews is an important task for online reputation management. The dataset of customer complaints - emotionally charged texts which are very similar to reviews and include descriptions of problems customers experienced with certain businesses – is presented. It contains 2746 complaints about banks and provides clear ground truth, based on available factual knowledge about the financial domain. Among them, 400 texts were manually tagged. Initial experiments were performed in order to explore the links between implicit cues of the rhetoric structure of texts and the validity of arguments, and also how truthful/deceptive are these texts.

CCS CONCEPTS

• Information systems → Information retrieval → Document representation → Content analysis and feature selection • Information systems → Information retrieval → Retrieval tasks and goals → Clustering and classification; Document filtering; Information extraction

KEYWORDS

Customer Complaints, Rhetorical Structure Theory, Discourse Analysis, Deception Detection, Fake Reviews

ACM Reference format:

Dina Pisarevskaya, Boris Galitsky, Jay Taylor and Andrey Ozerov. 2019. An Anatomy of a Lie:: Discourse Patterns in Customer Complaints Deception Dataset. In *Proceedings of WWW '19: The Web Conference (WWW '19), May 13, 2019, San Francisco, USA*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3308560.3316468

WWW '19, May 13, 2019, San Francisco, USA

ACM ISBN 978-1-4503-6675-5/19/05.
https://doi.org/10.1145/3308560.3316468

1 Introduction

It has been discovered that a lot of forms of human intellectual and communication activity are associated with certain discourse structures. Rhetorical Structure Theory (RST) [8] is a good means to express correlation between such form of activity and its representation in how associated thoughts are organized in text. Rhetorical Structure Theory present a hierarchical, connected structure of a text as a Discourse Tree, with rhetorical relations between the parts of it. The smallest text spans are called elementary discourse units (EDUs). In communicative discourse trees (CDTs), the labels for communicative actions (CAs) (VerbNet expressions for verbs) are added to the discourse tree edges to show which speech acts are attached to which rhetoric relations; this structure helps to understand argumentation [5, 26].

Argumentation needs a certain combination of rhetorical relations of *Elaboration*, *Contrast*, *Cause* and *Attribution* [18]. Persuasiveness relies on certain structures linking *Elaboration*, *Attribution* and *Condition* [19]. Explanation needs to rely on certain chains of *Elaboration* relations plus *Explanation* and *Cause*, and a rhetorical agreement between a question and an answer is based on certain mappings between the rhetorical relations of *Contrast*, *Cause*, *Attribution* and *Condition* between the former and the latter [23, 27]. Discourse trees turned out to be helpful to form a dialogue and to build dialogue from text, to better understand the structure of texts.

In this paper, we study rhetoric structure correlated with certain forms of verbal activity, namely we focus on deception in texts such as reviews and complaints. Automated detection of fake reviews is important for online reputation management tasks. Natural Language Processing tools, that could distinguish truthful and reliable reviews from deceptive reviews, could be important for a broad spectrum of applications of recommendation and security systems, for wide range of products and services. Research on automated deception detection in written texts is focused on classifying if a narrative is truthful or deceptive. The main difficulty is to detect deception where factual knowledge is not available to a degree sufficient to computationally establish the truth. This situation is typical in the every day life in the real world, from intuitive choice of product based on reviews to judges' verdicts: it is impossible to establish the truth based on known facts so decision are based on implicit cues such as the way people explain what they have done and provide arguments.

Detecting misrepresentation in writing, it is impossible to differentiate between different categories of writers. Professional writers are frequently good at misrepresenting, and they do not include cues for what might be a lie. Conversely, a content written by non-professional writers is often authentic in how it indicates the thought patterns of the writer where the traces of a lie and hints for how it is motivated can be found. Here, a corpora with defined ground truth are needed for classification tasks solving and exploring the links between implicit cues of rhetoric structure of texts and how truthful/deceptive are these texts [25].

2 Example of Misrepresentation in User-Generated Content

Regarding possible misrepresentation in a user-generated content, the following example can be provided:

> "I have accounts with them for almost 10 years, I hated it their customer service! Worst one ever. I don't know what's their problems, I'm not recommending their services and banking to anybody, I stopped using their credit cards already! The only reason I can't close my accounts with them, it could drop my credit score. I will not close my credit cards, but I'm not definitely using them so they can't make money from on us! I just had conversation with a supervisor from California called Steve he and his representative didn't even understand my situation, which was not common at all, basically didn't want to help me!"

The author of this complaint does not provide a single argument backing up his claim. And the author's statement that his credit history can be negatively affected by his closing an account is a misrepresentation.

We show the text split into elementary discourse units as done by discourse parser [20]. What do we see in the discourse tree for this text? We show important (non-default) rhetorical relations in bold and highlight with italics the verbs with the role of communicative actions which are important addition to the rhetorical relations.

Algorithm 1: A communicative discourse tree for the user-generated text example

elaboration (LeftToRight)
  elaboration (LeftToRight)
    attribution (LeftToRight)
      TEXT:I have accounts with them for almost 10 years ,
      TEXT:I *hated* it their customer service !
    TEXT:Worst one ever .
  elaboration (LeftToRight)
    elaboration (LeftToRight)
      explanation (LeftToRight)
        attribution (LeftToRight)
          cause (LeftToRight)
            attribution (RightToLeft)
              TEXT:I do not know
              TEXT:what is their problems ,
            TEXT:I 'm not *recommending* their services and banking to anybody ,
          TEXT:I *stopped using* their credit cards already !
        attribution (RightToLeft)
          <u>*TEXT:The only reason I can not close my accounts with them ,*</u>
          <u>*TEXT:it could drop my credit score .*</u>
      contrast (RightToLeft)
        TEXT:I will not close my credit cards ,
        enablement (LeftToRight)
          TEXT:but I 'm not definitely using them
          TEXT:so they can not make money from on us !
    elaboration (LeftToRight)
      TEXT:I just had conversation
      same-unit
        elaboration (LeftToRight)
          TEXT:with a supervisor from California called Steve , he and his representative did not even understand my situation ,
          TEXT:which was not common at all ,
        TEXT:basically did not want to help me !

There is an unusual chain of rhetorical relations explanation-attribution-cause-attribution-attribution which is a suspicious explanation pattern on its own. Unsurprisingly, the atom statement for the last attribution (which is the basis of this explanation, highlighted with underlined italics in Ex. 1) turns out to be false.

3 Example of Misrepresentation in Professional Writing

In our first example, the objective of the author is to attack a claim that the Syrian government used chemical weapon in the spring of 2018 (Fig. 1). An acceptable proof would be to share a certain observation, associated from the standpoint of peers, with the absence of a chemical attack. For example, if it is possible to demonstrate that the time of the alleged chemical attack coincided with the time of a very strong rain, that would be a convincing way to attack this claim. However, since no such observation was identified, the source, Russia Today, resorted to plotting a complex mental states expressing how the claim was communicated, which agents reacted which way for this communication. It is rather hard to verify most statements about the mental states of involved parties. We show the text split into EDUs as done by the discourse parser:

[Whatever the Douma residents ,][who had first-hand experience of the shooting of the water][dousing after chemical attack video ,][have to say ,][their words simply do not fit into the narrative][allowed in the West ,][analysts told RT .] [Footage of screaming bewildered civilians and children][being doused with water ,][presumably to decontaminate them ,][was a key part in convincing Western audiences][that a chemical attack happened in Douma .] [Russia brought the people][seen in the video][to Brussels ,][where they told anyone][interested in listening][that the scene was staged .] [Their testimonies , however , were swiftly branded as bizarre and underwhelming and even an obscene masquerade][staged by Russians .] [They refuse to see this as evidence ,][obviously pending][what the OPCW team is going to come up with in Douma], [Middle East expert Ammar Waqqaf said in an interview with RT .] [The alleged chemical incident ,][without any investigation , has already become a solid fact in the West ,][which the US , Britain and France based their retaliatory strike on .]

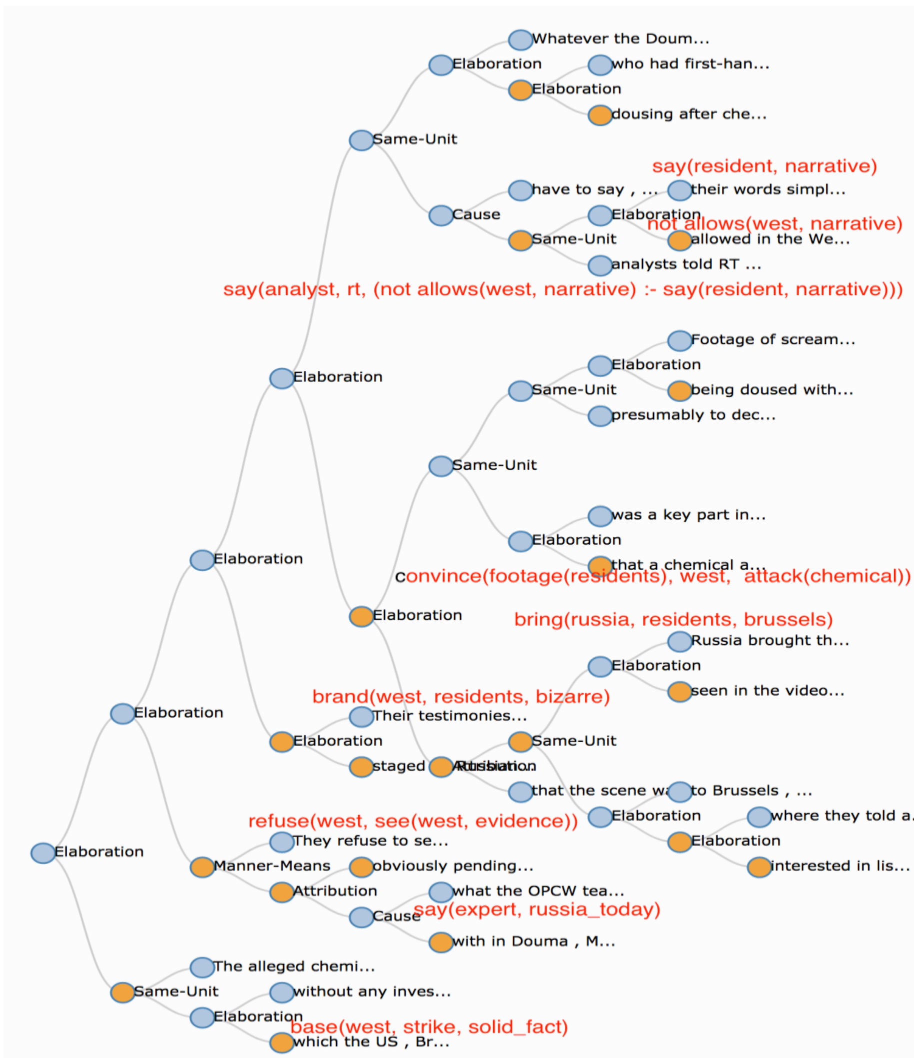

Figure 1: A DT for the chemical attack claim. An author attempts to substitute a desired valid argumentation chain by a fairly sophisticated mental states expressed by communicative actions.

This article (Fig. 1) does not really find counter-evidence for the claim of the chemical attack it attempts to defeat. Instead, the text says that the opponents are not interested in observing this counter-evidence. The main statement of this article is that a certain agent "disallows" a particular kind of evidence attacking the main claim, rather than providing and backing up this evidence. Instead of defeating a chemical attack claim, the article builds a complex mental states conflict between the residents, Russian agents taking them to Brussels, the West and a Middle East expert.

4 Background and Related Works on Deception Datasets

Deceptive product reviews can be referred to as deceptive opinion spam: fictitious opinions that have been deliberately written to sound authentic, in order to deceive the reader [13]. Spammers write fake reviews to promote or demote target products. They are deliberately written to sound authentic, and it is difficult to recognize them manually: human average accuracy is merely 57.3% [13].

Automated deception detection for reviews faces the lack of gold standard corpora with verified examples of deceptive uses of language. Besides this, intentionally written (e.g. by crowdsourcing) texts are distinct from genuinely produced texts. Hence, such artificial texts classified as deceptive by human annotators are not necessarily totally deceptive.

The release of two gold-standard datasets (available at http://myleott.com/) allowed for applying supervised learning methods, taking stylistic, syntactic and lexical features into consideration [12, 13, 2, 3]. Hotels reviews were chosen for the datasets, because it was suggested that deception rates among travel reviews is reasonably small. The latter dataset includes, among other reviews, crowdsourced generation of deceptive reviews. It contains 400 truthful positive reviews from TripAdvisor; 400 deceptive positive reviews from Mechanical Turk; 400 truthful negative reviews from reviews websites; 400 deceptive negative reviews from Mechanical Turk.

Later researchers tried to overcome the lack of large realistic datasets on different topics and domains. For example, Yao et al. [17] apply a data collection method based on social network analysis to quickly identify deceptive and truthful online reviews from Amazon. The dataset contains more than 10,000 deceptive reviews in diverse product domains.

The problem of the mentioned above gold standard datasets is that the fake reviews were not taken from genuinely written ordinary reviews and manually classified as fake. Instead, they were written on demand by the Amazon Mechanical Turk workers, hence they are not indicative of deception [10]. However, they are accepted as gold standard datasets for this research field. Rules used in [13] to create ground truth datasets were also used in later projects, such as in [6].

The real-life Amazon dataset [7] contains reviews from Amazon.com (crawled in 2006). It is large and covers a very wide range of products. It was used, for example, in Sun et al. [16], namely, three domains: Consumer Electronics, Software, and Sports. The metadata in this dataset provides only helpfulness votes of the reviews.

In cases where there was no certain knowledge of the ground truth, different ways to collect reviews corpora, relying on other features, were used. For example, in [4] the DeRev corpus of books reviews, originally posted on Amazon, was collected using definite pre-defined deception clues. Book reviews in the corpus are marked as clearly fake, possibly fake, and possibly genuine. The corpus is constituted by 6,819 instances whose 236 were labeled with the higher degree of confidence and are considered as the gold standard.

In [14], two publicly available Yelp datasets were presented. They are labeled with respect to the Yelp's classification in recommended and not recommended reviews. Mukherjee et al. [9] found that the Yelp spam filter primarily relies on linguistic, behavioral, and social networking features. Classification provided by Yelp has been also used in many previous works before as a ground truth, where recommended reviews correspond to genuine reviews, and not recommended reviews correspond to fake ones, so these labels can be trusted. The YelpNYC dataset contains reviews of restaurants located in New York City (359,052 reviews; 10,27% are fake); the Zip dataset is larger, since it contains businesses located in contiguous regions of the U.S. (608,598 reviews; 13,22% are fake).

Big Amazon dataset is annotated with compliant/non-compliant labels. It has many different topics: from electronics and books to office products (https://s3.amazonaws.com/amazonreviewspds/readme.html). It contains labels about star rating, helpful vote, total votes, verified purchase, that could be used for making decisions.

Hence, the existing recent datasets rely on external factors provided by their source, such as review's rating, number of votes, social networking features of review's author, metadata features etc. They are not annotated manually. So, despite the presence of different corpora, lack of corpora with exact ground truth can be understood as a bottleneck in deception detection of online texts.

5 Dataset Description

We introduce the dataset of customer complaints – emotionally charged texts which include descriptions of problems they experienced with certain businesses. The dataset is freely available [24].

Raw complaints were collected from PlanetFeedback.com for a number of banks submitted in 2006-2010. The dataset consists of 2,746 complaints totally. 400 complaints were manually tagged with respect to the parameters related to argumentation and validity of text: perceived complaint validity, argumentation validity, presence of specific argumentation patterns, and detectable misrepresentation. Here, validity of information is connected with validity of arguments. The dataset contains texts with direct truth confirmation based on manual annotation. It contains authentic data: both truthful and deceptive reviews were taken from spontaneously written customers' texts. Among the manually annotated 400 complaints, 163 are invalid and 237 are valid.

The initial set of 80 complaints was tagged by the authors of the paper as experts. After that, three annotators worked with this dataset, having a set of definitions and applying them. Then precision and recall were measured by matching the tags done by the authors as the 'gold standard', after that the set of definitions was edited and elaborated. In the further work, the Krippendorff's alpha measure (for three annotators) was applied as inter-annotator agreement measurement, and it exceeds 80%. As it is possible to know, retrospectively and based on facts, the established ground truth, we suggest that the annotators can find out, with high confidence, what information in texts is deceptive. So the dataset would provide ground truth.

The rest 2,346 complains were auto-tagged based on the model trained on this 400 set. After that they have also been partially manually evaluated, so that the accuracy of auto tagging exceeds 75%.

Our dataset includes more complaints with intense argumentation in comparison with other argument mining datasets, such as [15, 1, 11]. For a given topic such as insufficient funds fee, this dataset provides many distinct ways of argumentation that this fee is unfair. Authors attempt to provide as strong argumentation as possible to back up their claims and strengthen their case.

If a complaint is not truthful it is usually invalid: either a customer complains out of a bad mood or wants to get compensation. However, if the complaint is truthful, it can still easily be invalid, especially when arguments are flawed. When an untruthful complaint has valid argumentation patterns, it is hard for an annotator to properly assign it as valid or invalid, without the guidelines. So, according to the guidelines for the manual tagging of the dataset, a complaint was considered as valid if a judge believed that the main complaint claim is truthful under the assumption that a complainant is making truthful statement. Valid complaint needs to include proper discourse and acceptable argumentation patterns.

Following this approach, a complaint is marked as truthful if a judge cannot defeat it, using commonsense knowledge, available factual knowledge about a domain or implicit, indirect cues. Inconsistencies detected by a judge also indicate that the complaint author is deceiving. Mentioning multiple unusual, very rarely occurring claims also indicate that the complaint author is deceiving. The judge does not have to be able to prove that the complainant is lying: judge's intuition is sufficient to tag a complaint as untruthful. We suggest that one can provide a valid argumentation and also provide a false statement in a single sentence: "Rule is like this <correct rule> and I followed it, making <false statement>". Conversely, one can be truthful but provide an invalid argumentation pattern "I set this account for direct deposit and sent a check out of it <truthful statement>, as my HR manager suggested <should not have followed advice from not a specialist in banking>". Therefore validity (of argumentation patterns) and truthfulness are correlated.

Furthermore, customer complaints have much more significance for well-being of customers in comparison with customer reviews. Therefore, tagged customer complaints have much more importance associated with truth/deception than customer reviews. Since reviews are associated with opinions which can be random and complaints with customers doing their best to achieve their goals, both the truth and a lie is much more meaningful and serious in comparison with review datasets.

Complaints usually have a simple motivational structure, are written with a fixed purpose. Most complainants face a strong deviation between what they expected from a service, what they received and how it was communicated. Most complaint authors report incompetence, flawed policies, ignorance, indifference to customer needs from the customer service personnel. The authors are frequently exhausted communicative means available to them, confused, seeking recommendation from other users and advise others on avoiding particular financial service. The focus of a complaint is a proof that the proponent is right and the opponent is wrong, resolution proposal and a desired outcome.

Complaints reveal shady practice of banks during the financial crisis of 2007, such as manipulating an order of transactions to charge a highest possible amount of non-sufficient fund fees. Moreover, the most frequent topic is about banks attempts to communicate this practice as a necessity to process a wide amount of checks. That's why the dataset collection is based on complaints of 2007.

Multiple argumentation patterns are used in complaints.

1. Deviation from what has happened from what was expected, according to common sense (most frequent). This pattern covers both valid and invalid argumentation (a valid pattern).
2. The second argumentation patterns cites the difference between what has been promised (advertised, communicated) and what has been received or actually occurred. It also mentions that the opponent does not play by the rules (valid).
3. A high number of complaints are explicitly saying that bank representatives are lying. Lying includes inconsistencies between the information provided by different bank agents, factual misrepresentation and careless promises (valid).
4. Complaints arise due to rudeness of bank agents and customer service personnel. Customers cite rudeness in both cases, when the opponent point is valid or not (and complaint and argumentation validity is tagged accordingly).
5. Complainants cite their needs as reasons bank should behave in certain ways. A popular argument is that since the government via taxpayers bailed out the banks, they should now favor the customers (invalid).

6 Communicative Discourse Trees to Represent Truthfulness in Text

Starting from the autumn of 2015, we became interested in the controversy about Theranos, the healthcare company that hoped to make a revolution in blood tests. Some sources including the *Wall Street Journal* started claiming that the company's conduct was fraudulent. The claims were made based on the whistleblowing of employees who left Theranos. At some point FDA got involved, and as the case develops, we were improving our deception detection techniques while keeping an eye on Theranos' story. As we scraped discussions about Theranos back in 2016 from the website, the audience believed that the case was initiated by Theranos competitors who felt jealous about the proposed efficiency of the blood test technique promised by Theranos. However, our analysis showed that Theranos was misrepresenting and our findings supported the criminal case against Theranos, which led to the massive fraud verdict. SEC says that Theranos CEO Elizabeth Holmes raised more than $700 million from investors "through an elaborate, years-long fraud" in which she exaggerated or made false statements about the company's technology and finances.

We now build an example of communicative discourse tree (CDT) for the Theranos attack on Wall Street Journal (WSJ) acquisition ([Fig. 2]): "*It is not unusual for disgruntled and terminated employees in the heavily regulated health care industry to file complaints in an effort to retaliate against employers for termination of employment. Regulatory agencies have a process for evaluating complaints, many of which are not substantiated. Theranos trusts its regulators to properly investigate any complaints.*"

To show the structure of a deception, discourse relations are necessary but insufficient, and speech acts are necessary but insufficient as well. For the paragraph above, we need to know the discourse structure of interactions between agents, and what kinds of interactions they are.

file(employee, complaint) is elaborated by retaliate(employee, employer), and

evaluate(regulation, complaints) is elaborated by *trust(Theranos, regulators, complainants).* Also, the top link in turn is elaborated by the bottom link. Once we involve the definitions of the verbs for these four communicative actions, the inconsistency is revealed.

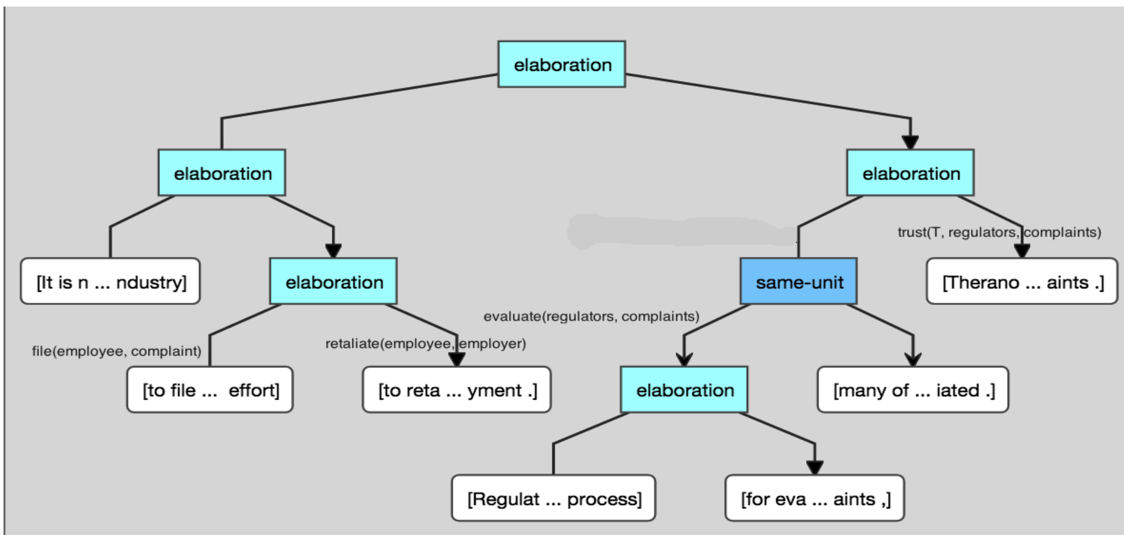

Figure 2: CAs as labels for rhetoric relations helps to identify a text apart from a heated discussion

From the commonsense reasoning standpoint, Theranos, the company, has two choices to confirm the argument that *his tests are valid*:

1) Conduct independent investigation, comparing their results with the peers, opening the data to the public, confirming that their analysis results are correct.
2) Defeat the argument by its opponent that their testing results are invalid, and providing support for the claim that their opponent is wrong.

Obviously, the former argument is much stronger, and we know, that usually the latter argument is chosen when the agent believes that the former argument is too hard to implement. On one hand, the reader might agree with Theranos that WSJ should have provided more evidence for its accusations against the company. On the other hand, the reader perhaps disliked the fact that Theranos selects the latter argument type (2) above, and therefore the company position is fairly weak.

The authors believe that Theranos' argument is weak because the company tries to refute the opponent's allegation concerning the complaints about Theranos's services from clients. We believe that Theranos' demand for evidence by inviting WSJ to disclose the sources and the nature of the complaints is weak. A claim is that a third-party (independent investigative agent) would be more reasonable and conclusive. However, some readers might believe that the company's argument (burden of proof evasion) is logical and valid.

7 Evaluation

In our evaluation we used the following pipelines:

Communicative Discourse Tree Construction. Just two RST parsers constructing discourse tree (DT) from paragraphs of text are available at the moment. We used the tool provided by [20, 21]. We then build CDT involving VerbNet.

Nearest Neighbor learning. To predict the label of the text, once the complete DT is built, one needs to compute its similarity with DTs for the positive class and verify that it is lower than similarity to the set of DTs for its negative class. Similarity between CDT's is defined by means of maximal common sub-DTs. Definitions of labeled graphs and domination relation on them used for construction of this operation can be found, e.g., in [22].

SVM Tree Kernel learning. A DT can be represented by a vector of integer counts of each sub-tree type (without taking into account its ancestors). For EDUs as labels for terminal nodes only the phrase structure is retained: we suppose to label the terminal nodes with the sequence of phrase types instead of parse tree fragments. For the evaluation purpose Tree Kernel builder tool [23] was used. After that, we applied the further set of more complex experiments. For all texts, we use CDT-kernel learning approach. We combined Stanford NLP parsing, coreference resolution tool, entity extraction, CDT construction (based on automated discourse parser as in [20, 21]) and Tree Kernel builder into one system that is presented in [25].

For the initial and automatically derived datasets, we show (in bold) the accuracies of training row and testing, averaging through 5x cross-validation. For the bottom three datasets, we tested the same SVM Tree Kernel model trained on our dataset. We demonstrate its universality, showing its applicability to texts of various nature such as consumer reviews. For genuine reviews, only 380 cases of deception were detected which were false positives, assuming that review writers do not lie (Table 1).

Table 1: Datasets, evaluation settings, accuracies for deception detection initial experiments

Dataset	Deception	No deception	P	R	F1
Manually tagged complaints	163	237	91	85	**88**
			83	81	82
Automatically tagged based on initial classifier	1132	1615	78	75	**76**
			69	71	70
Genuine reviews	380	3420	83	100	91
Fake reviews	414	286	100	59	74

Here, for reviews datasets we use the dataset presented in [12,13], in order to compare two following assessment frameworks:

1) The framework [12] which is trained on consumer reviews and tested on similar dataset;
2) Our deception recognition framework which we train once on our own dataset and tested on texts of various kinds, such as reviews.

We achieved the performance between 74 and 91% which is not as high as in [12,13] but by the universal text classification system. Hence we expect it to detect deception in other text datasets with acceptable accuracy to assure a resultant decision support system is usable.

In conclusion, we mention that our dataset is in the initial stage and is still being developed. In the future studies, the whole complaints dataset should be manually annotated and used for model training and new experiments, it could possibly help in results improvement. We are also going to run our detector again the business communication dataset from the real word, for evaluation purposes. We also plan to run experiments with other machine leaning methods. We also suggest that precision improvement (reducing the number of false positives) is mostly important for deception detection task, so we will implement further steps to improve precision. After that, we could also develop different methods of the customer complaints dataset extension.

Both truthfulness and validity are recognized reasonably well which is a value for Customer Relation Management systems and could be useful in different e-commerce tasks that are based on online review analysis. The dataset could be used for different machine learning models training that could help detect if reviews or other online texts of similar genres are truthful or deceptive, based on their content features.

REFERENCES

[1] Abbott, R., Ecker, B., Anand, P., and Walker, M.A. (2016). “Internet argument corpus 2.0: An sql schema for dialogic social media and the corpora to go with it”. Language Resources and Evaluation Conference.
[2] Feng, S., Banerjee, R., and Choi, Y. (2012a). “Syntactic stylometry for deception detection”. ACL 12, Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: 171-175.
[3] Feng, S., Xing, L., Gogar, A., and Choi, Y. (2012b). “Distributional footprints of deceptive product reviews”. Sixth International AAAI Conference on Weblogs and Social Media. 98-105.
[4] Fornaciari, T. and Poesio, M. (2014). “Identifying fake Amazon reviews as learning from crowds”. Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics: 279-287.
[5] Galitsky, B., Ilvovsky, D., and Pisarevskaya, D. (2018). “Argumentation in text: Discourse structure matters”. CICLing 2018 (unpublished).
[6] Hai, Z., Zhao, P., Cheng, P., Yang, P., Li, X.-L., and Li, G. (2016). “Deceptive review spam detection via exploiting task relatedness and unlabeled data”. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing: 1817-1826.
[7] Jindal, N. and Liu B. (2008). “Opinion spam and analysis”. Proceedings of International Conference on Web Search and Data Mining (WSDM-2008).
[8] Mann, W.C. and Thompson S.A. (1987). “Rhetorical Structure Theory: Description and Construction of Text Structures”. Springer Netherlands, Dordrecht.
[9] Mukherjee, A., Venkataraman, V., Liu, B., and Glance, N. (2013a). “What Yelp fake review filter might be doing?” Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media.
[10] Mukherjee, A., Venkataraman, V., Liu, B., and Glance, N. (2013b). “Fake review detection: classification and analysis of real and pseudo reviews”. Tech. rep. uic-cs-2013-03. University of Illinois at Chicago.
[11] Oraby, S., Reed, L., Compton, R., Riloff, E., Walker, M., and Whittaker, S. (2015). “And that’s a fact: Distinguishing factual and emotional argumentation in online dialogue”. NAACL HLT 2015 2nd Workshop on Argumentation Mining.
[12] Ott, M., Cardie, C., and Hancock, J.T. (2013). “Negative deceptive opinion spam”. NAACLHLT 2013, Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: 497-501.
[13] Ott, M., Choi, Y., Cardie, C., and Hancock, J.T. (2011). “Finding deceptive opinion spam by any stretch of the imagination”. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Volume 1: 309-319.
[14] Rayana, S. and Akoglu, L. (2015). “Collective opinion spam detection: Bridging review networks and metadata”. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM: 985-994.
[15] Stab, S. and Gurevych, I. (2017). “Recognizing insufficiently supported arguments in argumentative essays”. Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2017): 980-990.
[16] Sun, C., Du, Q., and Tian, G. (2016). “Exploiting product related review features for fake review detection”. Mathematical Problems in Engineering.
[17] Yao, W., Dai, Z., Huang, R., and Caverlee, J. (2017). “Online deception detection refueled by real world data collection”. Proceedings of Recent Advances in Natural Language Processing: 793-802.
[18] Grasso, F. (2003). “Characterizing Rhetoric Argumentation”. PhD Thesis Heriot-Watt University.
[19] Galitsky, B., Ilvovsky, D., and Makhalova, T. (2019). “Enabling a Bot with Understanding Argumentation and Providing Arguments”. Developing Enterprise Chatbots, Springer Switzerland.
[20] Joty, S., Carenini, G., Ng, R.T., and Mehdad, Y. (2013). “Combining intra-and multisentential rhetorical parsing for document-level discourse analysis”. ACL (1): 486-496.
[21] Surdeanu, M., Hicks, T., and Valenzuela-Escarcega, M.A. (2015). “Two Practical Rhetorical Structure Theory Parsers”. NAACL HLT.

[22] Ganter, B. and Kuznetsov, S.O. (2001). "Pattern Structures and Their Projections". ICCS, Lecture Notes in Artificial Intelligence (Springer), Vol. 2120: 129-142.

[23] Galitsky, B. (2019). "Rhetorical Agreement: Maintaining Cohesive Conversations". Developing Enterprise Chatbots, Springer – Cham, Switzerland.

[24] Github (2019). Ultimate Deception Dataset https://github.com/bgalitsky/relevance-based-on-parse-trees/blob/master/examples/ultimateDeceptionAutoTagged.csv.zip.

[25] Galitsky, B. (2019). "A Content Management System for Chatbots". Developing Enterprise Chatbots. Springer Switzerland.

[26] Galitsky, B. and Kuznetsov, S.O. (2008). "Learning communicative actions of conflicting human agents". J. Exp. Theor. Artif. Intell, 20(4): 277–317.

[27] Galitsky, B., Ilvovsky, D., and Kuznetsov, S.O. (2018). "Detecting logical argumentation in text via communicative discourse tree". J. Exp. Theor. Artif. Intell., 30(5): 1-27.